

What is claimed is:

1. A method for using an audio input from a telephony device to perform an action on an Internet protocol ("IP") network, the method comprising:
 - providing a telephony interface module;
 - receiving at the telephony interface module from the telephony device a first packet signal conforming to a telephony packet protocol and having an audio portion;
 - receiving at the telephony interface module from a second module in communication with the telephony interface module (i) a second packet signal conforming to an IP, the second packet signal having an audio portion and (ii) a command;
 - routing the first packet signal in accordance with the received command;
 - converting, in the telephony interface module, the second packet signal to a third packet signal conforming to a telephony packet protocol and including an audio portion; and
 - transmitting the third packet signal to the telephony device.
2. The method of claim 1 further comprising:
 - routing the first packet signal to a navigation module in communication with the telephony interface module;
 - converting, in the navigation module the audio portion of the first packet signal to a text equivalent signal;
 - converting, in the telephony interface module, the text equivalent signal to an IP network command signal; and
 - using the IP network command signal to retrieve a document from the IP network.
3. The method of claim 2 wherein the retrieved document is a voice XML document from the Internet.

4 telephony device over the PSTN and b) convert the circuit-switched signal to a telephony packet
5 protocol signal having an audio portion;
6 an Internet protocol ("IP") network;
7 an audio browser in communication with the telephony gateway to receive the telephony
8 packet protocol signal and in communication with the IP network..

1 11. The system of claim 10 wherein the audio browser further comprises:

2 a voice XML browser;
3 a navigation module;
4 a content retrieval module; and
5 a telephony interface module.

1 12. The system of claim 10 further comprising web cache.

1 13. The system of claim 11 wherein the navigation module further comprises one of speech
2 recognition module and touch tone (DTMF) recognition module.

1 14. The system of claim 11 wherein the content retrieval module further comprises one of
2 text-to-speech module and streaming media module.